POLICY RESEARCH WORKING PAPER 2131

Efficiency Wage and Union Effects in Labor Demand and Wage Structure in Mexico

An Application of Quantile Analysis

William F. Maloney
Eduardo Pontual Ribeiro

In this study of Mexican firms, unions appear to bargain principally about employment rather than wages and firms appear to pay efficiency wages above market clearing to reduce turnover. Since minimum wages are not binding, whatever labor market segmentation is observed arises endogenously, and is not due to union- or government-induced distortions.

The World Bank
Latin America and the Caribbean Region
Poverty Reduction and Economic Management Sector Unit
May 1999

Summary findings

Applying quantile analysis to detailed firm-level data from Mexico, Maloney and Ribeiro study determinants of demand and wages for two classes of labor.

Unions appear to have a strong impact on how much unskilled labor is employed but not on wages. This suggests an extreme example of "efficient bargaining" rather than the more common "monopoly union" behavior. The impact on productivity is, by definition, negative, but unions could also be said to be forcing firms to use "appropriate technology" (less capital and more workers), increasing the total amount of labor employed in the economy. The only impact on wages appears for the tenth (lowest) quantile of unskilled workers, suggesting that unions prevent workers from being paid too far below the median for their skill level.

Maloney and Ribeiro identify significant efficiency wage effects where firms pay above market clearing to prevent labor turnover both in labor demand and in the wage equations. Since minimum wages are not binding and the union impact on wages is small, this suggests that whatever segmentation exists emerges endogenously and is not due to union- or government-induced distortions.

Maloney and Ribeiro offer the first use of quantile analysis to analyze labor demand at the firm level, and one of the first uses of correct standard errors in two-stage least-squares quantile regression.

This paper — a product of the Poverty Reduction and Economic Management Sector Unit, Latin America and the Caribbean Region — is part of a larger effort in the region to understand the functioning of developing country labor markets. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington DC 20433. Please contact Tania Gomez, room I8-102, telephone 202-473-2127, fax 202-522-2119, Internet address tgomez@worldbank.org. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/html/dec/Publications/Workpapers/home.html. The authors may be contacted at wmaloney@worldbank.org or ribeiro@vortex.ufrgs.br. May 1999. (43 pages)

Efficiency Wage and Union Effects in Labor Demand and Wage Structure in Mexico: An Application of Quantile Analysis

William F. Maloney
The World Bank

Eduardo Pontual Ribeiro*
Universidade Federal do Rio Grande do Sul

* We thank Carlos Arango, Omar Arias, Gustavo Gonzaga, Daniel Hamermesh and Lyn Squire for helpful comments and to Tania Gomez for vital assistance. We are also grateful to the Mexican National Institute of Statistics, Geography and Information (INEGI) for the use of the data. INEGI is in no way responsible of any incorrect manipulation of the data or erroneous conclusions drawn from it. The views expressed here are those of the authors only, and should not be associated with the World Bank and its member countries. Contacts: wmaloney@worldbank.org, ribeiro@vortex.ufrgs.br

I. Introduction.

The evidence to date suggests that both union power and efficiency wage behavior may have large effects on the structure and dynamics of labor markets. The literature documenting their effect on wages, in particular, is vast. However, as Blanchflower et. al. (1991) note, the impact on employment of unions has received relatively little study, and that of efficiency wages has attracted even less. Further, as Nickell and Wadhwani (1991) argue, the existence of both phenomena simultaneously complicates efforts to distinguish between competing models of union behavior.[1] Their work and that of Hendricks and Kahn (1991) are among the very few that analyze employment determination in the presence of efficiency wages and two types of union bargaining: the "right to manage" (RTM) or "monopoly union" type where unions attempt to set the wage but let firms choose the level of employment, and the "efficient bargaining" (EB) type where unions bargain over both (Oswald 1991 and Layard and Nickell 1990). Using a large panel from the UK manufacturing sector, Nickell and Wadhwani find no evidence of union influence on employment and mixed evidence of efficiency wage effects. Hendricks and Kahn find EB effects in their study of the demand for police in the US.

This paper builds on this work in two ways. First, it approaches these issues using quantile analysis that more completely characterize the distributions of wages and labor demanded than can be done using the conditional mean based linear regression approaches (OLS, 2SLS) that are standard. Despite assumptions of identical agents, in fact the samples may show substantial heterogeneity and large differences in the impacts of regressors across

[1] For discussions of the theory of efficiency wages see Stiglitz (1974), Krueger and Summers (1988), Phelps (1994) and Weiss (1990). For a review of the literature on union impacts, see Lewis (1986).

quantiles. For example, union power might be expected to impinge more strongly among those who receive a relatively high wage given their human capital. But, alternatively, unions may put a floor under the wage such that workers whose productivity is very low given their nominal human capital and leave the rest of the wage distribution market determined. Efficiency wage effects would probably also be expected to be most prevalent among those paid "above the market" given their human capital. The results show the power of this technique to uncover differences that would ordinarily go undetected. The paper offers the first use of quantile analysis in firm level analysis of labor demand and the first use of correct standard errors in two stage least square quantiles.

Second, it analyses Mexico, a country with unique institutional, economic and political characteristics that make it an important case study to add to the literature for two reasons. First, we may observe union behavior that, although theoretically plausible is contrary to that commonly found. The manufacturing sector studied here is heavily unionized: 18% of firms have no union representation, and the rest have a median unionization rate of 70%. However, the massive Labor Congress (CT) which embraces the Confederation of Mexican Workers (CTM), the Revolutionary Federation of Workers and Peasants (CROC), the Federation of Government Workers (FSTSE) and roughly 38 other labor organizations has had a longstanding and close relationship with the governing Revolutionary Institutional Party (PRI). Particularly since 1987 with the inception of the *Pacto Social*- a joint agreement of labor, business and the government to promote price stability- unions have closely coordinated wage demands with national stabilization objectives. This has made possible extreme downward flexibility of wages during recent

crises and may have curtailed union power along this dimension.[2] In addition, it can be argued that in the absence of unemployment insurance, employment enters more heavily in union objective functions than in the industrialized countries. The particular constraints and elements of union utility may give rise to EB outcomes with implications for the structure of wages, and the overall level of employment distinct from those generally anticipated. It also may permit a test of the assertion (see Nickell and Wadhwani, Layard and Nickell among others) that the wage-employment relation may slope upward.

Second, the Mexican labor market is well suited to testing for efficiency wage effects. Though a longstanding literature explains dualistic LDC labor markets by government or union interference in the wage setting process,[3] Mexico's minimum wage was not binding in the period we study (Bell 1997), and if unions focus primarily on employment, then the wage structure and whatever dualism is observed may be emerging endogenously through efficiency wage effects. The existence of a large non-unionized sector permits isolating such effects whose manifestations can sometimes be indistinguishable from the outcomes of union bargaining. Further, since the Mexican Constitution prohibits firing of workers except in extreme circumstances we are arguably testing for one particular variety of efficiency effect arising from the prevention of turnover (Stiglitz 1974).

The data set we work with is exceptionally rich. It permits conditioning on numerous dimensions of firm heterogeneity as well as offering some that may conceivably be associated with efficiency wage effects. It provides evidence on the dynamics of unionized firms and

[2] Though in November of 1997, the New Union of Workers (UNT, .7-1.5 million workers) split from the CTM largely over what was perceived to be excessive responsiveness to government initiatives, across the period analyzed here some analysts have seen a decline in union influence both within the PRI and overall. This discussion partly based on Collier and Collier (1991) and Brooks and Cason (1998).

their behavior in the use of three inputs: unskilled and skilled labor and, to a lesser degree, capital. Disaggregating labor promises an improvement over the vast majority of efficiency bargain papers as we may avoid a composition bias in the demand for labor due, for example, to the substitutability of types of workers. Further, we can go beyond the standard (static) labor demand literature with worker types (e.g. Hamermesh, 1993 and references therein)[4] that has not allowed for the possibility of employment decisions occurring along a contract curve as in the EB solutions instead of the standard labor demand curve.

The paper is organized as follows. The following section presents the theoretical background and the testable implications. The third section discusses quantile analysis. The fourth presents the data set used in the paper and the fifth the empirical results. The last section concludes with a summary of results.

IIa. Analytics: (Overview)

Efficiency Wages:

The extensive literature on efficiency wages provides a rationale for firms to *voluntarily* pay wages above the market clearing level. One common variant of these models arises from the difficulty of monitoring individual workers and the lack of any penalty from being caught "shirking" - any activity, or lack thereof, that might be detrimental to the firm. If wages are market clearing, a worker fired for shirking can simply get another job at the same wage. However, if all firms pay higher than market clearing wages, unemployment will be

[3] See, for example, Harris and Todaro(1970) See Esfahani and Salehi Isfahani (1989) as an example of modeling LDC dualism in an efficiency wage context.
[4] Hamermesh also points out the clear advantages of using microdata and the dearth of such studies.

created in the economy that creates a disincentive to being laid off and hence to shirking. [5]

Since, in many Latin American countries, workers can be fired only with difficulty, the "turnover" variant of efficiency wage models is probably more appropriate: firms must invest resources in workers when they are hired, perhaps through training or through the process of recruitment, that will be lost if the worker leaves. Hence, it is worthwhile for firms to pay higher wages and raise the opportunity cost of leaving.

Interviews with Mexican entrepreneurs in the survey used here support this view. Roughly 30% stated that the resignation of recently trained workers was a problem. This is almost certainly an understatement for two reasons. First, "recently" may not capture the relevant period of return on the investment in the worker. Second, if the firm is already paying the optimal efficiency wage to prevent workers from leaving, it will not report excessive turnover as a problem. Of those reporting frequent resignations after training, 58% do something to raise the total well-being of the worker after training, 28% raise remuneration without promoting the worker, and 40% take measures that increases the wage of the worker, including promotions (see Appendix I).

The efficiency wage argument is particularly compelling in LDCs where firms may absorb a larger share of education costs due to poorly functioning education systems. Thus, firms will be very concerned about preventing workers they train from moving to another firm. In addition, in countries where self-employment (formal or informal) are considered desirable destinations, it is possible that workers enter formal salaried work to accumulate

[5] As Marquez and Ros (1990) noted, and has been confirmed by later studies, wages of similar workers rise with firm size, much as they do in industrialized countries. Further, Marquez (1990), Abuhadba and Romaguera (1993) and Schaffner (1998) find efficiency wage effects in the patterns of wage differentials that are strong and highly correlated among Chile, Venezuela, and Brazil and the U.S.. This suggests that the conditional wage dispersion (wages adjusted for human capital) and rigidities may be emerging endogenously and are not due to

skills and financial capital, and then quit to open their own business.

Both theories imply that the offers workers can get outside the firm (the outside wage), as well as the probability of being able to get a job at that wage (the hiring rate) should be important to determining the wage that is set in the firm, as well as to the quantity of labor hired.

Union Bargaining.[6]

The most common view postulates that unions maximize utility, which may be a function of both the wage received by union members and the level of employment, subject to a constraint representing combinations of the two that firms are willing to pay, the labor demand curve. In the "Right to Manage" view unions would identify the level of the wage that maximizes their utility, and firms simply set the level of employment.

However, if the firm is a monopolist or oligopolist and earns excess profits, then both unions and firms may be better off by coming to a bargain that pushes them off the labor demand curve. Figure 1 traces out a series of iso-profit curves- combinations of the wage and level of employment such that the firm earns the same level of profits. A lower curve implies a higher rate of profits. The apex of each curve is necessarily cut by the labor demand curve: the firm maximizes profits subject to any given wage, that is, it chooses the level of employment that puts it on the highest iso-profit curve possible. Point P represents one such point. As the wage rises or falls, the firm's optimal level of hiring traces out the labor demand curve, the locus of all apexes of iso-profit curves. Employment either below or above the profit maximizing level (100 at W_o) necessarily implies that the firm earns fewer profits and is

---

either government or union intervention.

[6] Graphs taken from and discussion based on Borjas (1996).

thus on a higher iso-profit curve. Therefore the iso-profit curves must slope downward on either side of the intersection with the labor demand curve.

As point M in figure 2 shows, a better deal for both workers and firms can be negotiated than that at "Right to Manage" equilibrium at point M. Here, the union's utility curve is tangent to the demand curve, but not to the iso-profit curve of the firms. Thus, the willingness of workers and firms to trade off employment for wages is not equal, and the equilibrium is not efficient. Two alternate and more efficient bargains where the two curves are tangent can easily be seen, both of them off the labor demand curve. First, at point R, Unions reach a higher level of utility, U_R compared to U_M while firms are earning the same level of profits. Alternately, at Q, unions are no worse off while firm profits are higher. Which bargain, R, Q or perhaps Q', where both are better off, are "efficient bargains" and lie on the contract curve. The contract curve is the set of efficient bargains ranging along the line PZ from P, where workers have no bargaining power and take the market wage W* and the firm takes all profits, π^*, to Z where π_Z represents the level of profits below which the firm would go out of business, and the union captures all of the monopoly rents. This iso-profit curve also suggests that the maximum wage workers could ever gain would be W_z and then only if it cares very little about employment.

These bargains along PZ, however, are clearly not efficient from a production point of view: at any bargain except P, more workers are being hired than the firm would hire in the absence of a union, E*. This "featherbedding" is a way of transferring firm profits to workers through the creation of unnecessary positions, rather than wages. The final equilibrium clearly depends then on the goals of the union as captured in the shape of its utility function, that jointly with the firm's iso-profit functions determines the contract curve, and the union's

relative bargaining strength, which determines the position of the final bargain along the contract curve.

Both union objectives and bargaining power in Mexico may be different from those in industrialized countries for a variety of reasons. First, like much of Latin America during the 1980's and early 90's, job growth has been slow relative to population growth. Second, as is the case with most of its neighbors, Mexico has no system of unemployment insurance and employment stability may be more highly valued than wages. Third, since the post-revolution inception of the Institutionalized Revolutionary Party (PRI) in 1929, the major unions have had a longstanding and close relationship with the government. Particularly since 1987 with the inception of the *Pacto*- a joint agreement of labor, business and the government to promote price stability- unions have closely coordinated wage demands with pacto guidelines. These factors taken together may lead to an emphasis on employment creation, relative to pushing up wages in the union utility function.

The next section details how empirically it is possible to determine whether the type of bargaining occurring as well as if efficiency wage effects are important.

IIb. Analytics (detail):

Broadly following Nickell and Wadhwani and Layard and Nickell we postulate a firm facing a downward sloping product inverse demand curve $d(\,)$ with shift term, σ. Its real revenue function

$$R(N,\Omega,e,\sigma) = F(N,\Omega,e)d(F(.),\sigma) \quad R_1 > 0, R_2 > 0, R_3 > 0$$

is a function of the labor it hires, N, the stock of other factors including capital, management ability, technology, Ω, and also efficiency wage effects on the productivity of labor, *e*. Among these is the ratio of the inside wage, *W*, to the expected alternative outside wage, $E(W_a)$. Firms bargain with a union whose utility

$$u = U(W, E(W_a), N) \quad U_1 > 0, U_2 < 0, U_3 > 0$$

depends on the wage, the expected outside wage, and employment. In the "right to manage" model the union bargains for a level of *W*, and lets the firm choose the level of employment. However, if the union cares about employment as well, then its utility is maximized over both *N* and *W* and the outcome is determined jointly in an "efficient bargain" with the firm. In this case, the firm moves off the demand curve it would face in the RTM scenario and onto the contract curve.

The result of a standard Nash bargaining model yields a system of equations, both for employment and the wage. The firm solution is a system of equations for labor demand and wages of an (implicit) form such as

$$N = N(W, E(W_a), Z_2, e, \theta_N)$$

$$W = W(E(W_a), Z_2, e, \theta_W)$$

that reflect the compound effects of the two utility functions, as well as union bargaining power over employment, θ_N and the wage, θ_w. Z_2 contains variables that determine the position of the labor demand relation, such as Ω and σ. The expected outside wage enters both through the union utility function, and efficiency wage effects.

Several empirically testable predictions derive from the Nickell and Wadhwani and Layard and Nickell framework which are testable with the Mexican data:

a. If unions bargain solely over the wage, then union power will be captured entirely in the wages paid by the firm and free-standing proxies for union power should have no effect in labor demand functions. Alternatively, if unions also bargain over the level of employment, the union proxy should enter positively in the demand equation. In the extreme case that unions do not bargain over the wage, but only employment, the union terms should be insignificant in the wage equation.

b. Since the workers' alternative, the expected outside real wage adjusted for the probability of getting a job, enters both in the firm's calculation of the optimal efficiency wage as well as the union utility function, its predicted sign and magnitude are ambiguous in cases where union power is present.[7] To avoid this problem, we will work with both unionized and non-unionized sectors to search for efficiency wage effects.

c. The sign of the employment/wage elasticity depends on whether unions have more power bargaining over employment or over wages.

d. As union power over employment determination θ rises, the elements of Z_2 (Ω, and σ) should lose influence in the labor demand equations.

We are particularly concerned with how these effects vary across a heterogeneous sample. Most obviously, if unskilled workers are represented by unions more than the skilled, we may observe different union and efficiency wage effects for each group. However the quantile analysis detailed below also allows us to investigate whether these factors impinge differently across the conditional distribution within these two samples.

[7] Nickell and Wadhwani argue that the appearance in the demand function of outside wages indicates the presence of efficiency wages unless unions both bargain over employment and more importantly, have a non-standard objective function, with the sign depending on the size of the standard employment-wage elasticity.

III. Empirical Methodology

Conditional mean regression estimators, such as Ordinary Least Squares, are traditionally used to estimate the relations such as those posited above. Minimizing the squared sum of errors allows estimating the values of the parameters that predict the mean of the dependent variable, conditional on a set of explanatory variables chosen. However, if the sample is not completely homogeneous, such techniques may hide differential effects of the regressors across the distribution that may be a critical part of the story being told. Further, if there are large outliers, or the distribution of the disturbances is non-normal, conditional mean estimators may be inefficient and often biased.

These concerns can be reduced somewhat by estimating the conditional median regression where half the errors lie above, and half below the fitted curve. Quantile analysis, introduced in Koenker and Bassett (1978), extends this analysis to estimating curves where approximately τ% of the errors will be negative and (100-τ)% of the errors will be positive.[8] If the errors are i.i.d., slicing the distribution at different quantile levels has little effect on parameter estimates and little information is lost in a single measure of the conditional central tendency, such as the parameters generated by OLS or median regression. However, figure 3 shows that asymmetries or heteroskedasticity in the distribution of errors may lead to substantially different estimates of the impact of the variables under study.

The problem of estimating an equation with endogenous explanatory variables under quantile analysis was addressed successfully by Powell (1983). A two stage method, where a least square regression is run on the first stage and median regression on the second as in 2SLS, was shown to generate consistent estimates with asymptotically normal distributions

under weaker assumptions than least squares. This special case of a two-stage quantile regression (2SRQ) was generalized for any quantile by Chen and Portnoy (1996).

In all the empirical work below, we present results of the quantile analysis at τ= 50 (the conditional median regression) completely, τ= 10 where 10% of the deviations lie below the estimated regression, and τ= 90 where 90% lie below. Appendix II presents the standard conditional mean regressions, whether OLS or 2SLS for reference. In all cases they are very close the median regression.

Correct Standard Errors for Two Stage Regression Quantiles

Standard errors estimates for regression quantiles have been studied in Buchinsky (1995) for models with exogenous regressors. Based on a Monte Carlo study, the author recommends the use of the *design matrix bootstrap*, as this method is valid under many forms of heterogeneity (heteroscedasticity), with a small reduction in efficiency in *iid* samples, compared to other methods. As in our model we cannot reject *apriori* heterogeneity (confirmed by LS based heteroscedasticity tests), so we choose this method to estimate the covariance matrix of the regression parameter vector.

The method amounts to sampling pairs (y_i^*, x_i^*) in a regression model $y_i = x_i'\beta + u_i$ to generate a pseudo sample of the data and obtaining an estimate b^* of β. The process is repeated B times and the B estimates of β are used to construct the covariance matrix. The pseudo sample can be of size n, the original sample size, as in this paper, and B should be large enough to guarantee a small sample variability of the covariance matrix. We chose

[8] The technique has generally been applied to estimating returns to education, (Buchinsky 1994).

B=200, based on the literature. The use of the design matrix bootstrap method for models with endogenous regressors can be argued for using the results of Freeman and Peters (1994) on bootstrapping 2SLS models and the analogy principle of estimation in Mansky (1988). In the present case, the covariance matrix for the labor demand equations were obtained sampling the triplets (y_i^*, x_i^*, z_i^*), where z_i is the vector of instruments, or exogenous variables in the system and x_i includes the endogenous explanatory variables. Both first and second stage regressions are then run to obtain the estimates of the parameter vector β for each of the B samples.

IV. Data:

We employ *the Encuesta Nacional de Empleo, Salarios, Tecnologia y Capacitacion* (ENESTYC), the National Survey of Training, carried out by the Mexican Official Statistics Institute (*Instituto Nacional de Estadistica, Geografia e Informatica*, INEGI) for the year 1992 which contains detailed information on firms specific variables relating to employment, technology, capital stock, etc. A 1995 Survey was also available that had the advantage of collecting data on share of the work force unionized at the firm level. However, it lacked information on the human capital of the work force and because the period it spanned contained the Tequila crisis in December 1994 and the beginning of the ensuing recession, we work primarily with what may be considered a more "normal" period of relative prosperity.

Variables:

Core Variables:

Wages and Employment: Following Roberts and Skoufias (1997) and others the wage and labor stock of skilled ($W_S$and N_S, respectively) and unskilled labor ($W_U$and N_U, respectively) are derived as weighted averages of subcategories within each. The weights for constructing the labor variables are the full wage (wage, social security and other non-wage benefits) per worker that capture the relative "marginal product" of each subclass. This generates a compound measure of "efficient units" of skilled or unskilled labor with the least productive subclass of labor as the numeraire in each.[9] The wage is then the total payments to the subclasses of labor divided by the labor measure, which, in practice is simply the wage of the numeraire subclass. The average schooling of the unskilled is about half of the skilled workers.

Value Added (Value Add.)*:* the value of total 1991 output minus the expenses in materials and energy in million Pesos.

Human Capital Variables:

Schooling (School and School2): Average years of schooling of the employed workers in each skill level in the firm, where the years of schooling were obtained from 7 levels.

Experience(Experience and Experience2): Average tenure in the firm of workers within each sub-class of labor.

[9] This approach is arguably preferable to simply assuming that each subclass of workers has identical productivity in the aggregation. In the skilled category are found directors (*directivos)*, Professionals (*profesionista)*, Technical workers (*tecnicos)*, Administrative Employees (*empleados administrativos)* and Supervisors (*supervisores)*. Among the Unskilled are professional workers *(obreros profesionales)*, specialists (*especializados)* and general workers (*general)*.